Review

Biomarkers in Upper Aerodigestive Tract Tumorigenesis: A Review

Dong M. Shin,1,2,3 Walter N. Hittelman, and Waun K. Hong

Departments of Thoracic/Head and Neck Medical Oncology [D. M. S., W. K. H.] and Clinical Investigation [W. N. H.], The University of Texas M. D. Anderson Cancer Center, Houston, Texas 77030

Abstract
Because therapeutic efforts such as surgery, radiotherapy, and chemotherapy have only marginally improved the 5-year survival rate from cancers of the upper aerodigestive tract (including head and neck and lung cancers) over the past 2 decades, chemoprevention has become an important strategy in reducing the rates of incidence and mortality of these cancers. However, chemoprevention trials have been hampered by serious feasibility problems; they require large numbers of subjects and long-term follow-up for accurate determination of cancer incidence and they are very costly. Because the use of intermediate end points would reduce the duration and costs of these studies, biomarkers that could serve as such end points may have become a subject of great interest. With the strengthening of the assumption that tumorigenesis is a multistep process of transformation from normal tissues to malignant lesions, there has been a great effort to examine each of these steps for genetic and/or phenotypic alterations that might be candidates for such biomarkers. These candidates include genomic markers, certain specific gene alterations, such as tumor suppressor genes, oncogenes, growth factors and their receptors, proliferation markers, and differentiation markers. In this review, we describe several genomic markers, including micronuclei, chromosomal alterations, and specific genetic markers, e.g., the ras gene family, erb B1, int2/htk-1, and p53 tumor suppressor gene. We also review the proliferation markers, including proliferating cell nuclear antigen, and squamous cell differentiation markers, including keratins, involucrin, and transglutaminase 1. These biomarker candidates have the potential to be important adjuncts to the development of new chemopreventive agents and to the rational design of future intervention trials. However, we cannot overemphasize that these markers need to be validated in clinical trials; only then can they replace cancer incidence as the sole end point for chemoprevention trials.

Introduction
Epithelial cancers of the upper aerodigestive tract are an increasingly important public health problem throughout the world. Despite improvements in surgery, radiotherapy, and chemotherapy over the past 2 decades, the 5-year survival rates for head and neck and lung cancers have improved only marginally. New research directions are clearly needed. Interest has been renewed in chemoprevention as a means of reducing the incidence and mortality of these cancers (1–3). Unfortunately, chemoprevention approaches have been hampered by serious feasibility problems associated with the conduct of randomized Phase III trials. Investigators have been forced to rely on cancer incidence as the study end point for determining preventive efficacy. This requires that chemoprevention trials have large numbers of subjects and long-term follow-up, measures which make the trials very costly. Therefore, there has recently been a great surge of interest in defining “biomarkers” associated with specific stages of the carcinogenic process, with the goal of using these markers as “intermediate end points” in chemoprevention trials (4, 5). This concept has been well described by Zelen (6), who stated that intermediate end points, such as biomarkers, would make prevention trials feasible.

Interest in systemic treatments that would prevent cancer in the aerodigestive tract, springs from the understanding that the whole epithelial lining of the tract shares both common carcinogenic exposure and the resulting increase in cancer risk. This entire epithelium was first described as mucosa “condemned” by carcinogens by Slaughter et al. (7) in 1953. They coined the term “field cancerization” to describe the diffuse histologic abnormalities and multifocal nature of squamous cell carcinomas of the head and neck. In the case of aerodigestive tract cancers, this hypothesis is supported clinically by the frequent association of tumors with premalignant lesions in the same field, e.g., oral leukoplakia (8) or bronchial metaplasia and/or dysplasia (9) and by the synchronous or metachronous development of multiple primary tumors (10).

Another concept of carcinogenesis in the aerodigestive tract is that of the multistep tumorigenic process (11). The driving force behind this process is thought to be genetic damage caused by continuous exposure to carcinogens, as evinced by an increased frequency of micronuclei in high-risk tissue and in premalignant lesions (12). Eventually, these genetic alterations give rise to phenotypic changes in the tissues, such as dysregulation of cell proliferation, differentiation, and cell loss pathways. These phenotypic alterations can be driven by alterations of certain specific genes such as tumor suppressor genes (e.g., p53 and rb), oncogenes (e.g., ras and myc), or the genes for growth...
The process of developing biomarkers from the laboratory to clinical application in humans requires the application of new insights to existing principles. The domain of biomarker identification falls primarily to basic scientists such as tumor biologists who are in seeking to apply markers for early detection of cancer, faced with three fundamental issues (13): (a) to provide a clear definition of the end point for which the putative index is a marker; (b) to identify the type of clinical specimen from which the marker can be measured; and (c) to establish an expected (i.e., normal or background) range of marker variability. For upper aerodigestive tract, on the other hand, Lippman et al. (4) proposed the following criteria for biomarkers in tobacco-related epithelial carcinogenesis: (a) that their expression in normal tissue be different from that in high-risk tissue; (b) that they can be detected even in small tissue specimens; (c) that they are expressed in a quantity or pattern that can be correlated with the stage of carcinogenesis; and (d) that preclinical or early clinical data indicate that the condition represented by the marker can be modulated by study agents.

A paradigm of carcinogenesis presupposes a multistage model featuring genetic or phenotypic markers for each stage. Tumor development is characterized by stepwise genetic changes from a normal to a malignant cell. Elegant models have been worked out by many investigators to define the multistep nature of the carcinogenesis process (14–16). In these models, cellular or genetic alterations may precede the occurrence of invasive malignancy and, if they do, can be used as intermediate end points of carcinogenesis. Biomarkers of these intermediate end points, then, can be defined as measurable markers of cellular or molecular genetic events associated with specific stages of cancer development. This definition indicates that the risk of carcinogenic transformation correlates with the quantitative degree and pattern of biomarker expression. Because the search for biomarkers useful in chemoprevention is just beginning, no individual biomarker or pattern of biomarkers has so far been validated through prevention trials as a conclusive predictor of risk.

To be a candidate biomarker in tumorigenesis, the marker should be altered not only in the malignant stages but also in the premalignant stages of cancer development. In this sense, activation of oncogenes and inactivation of tumor suppressor genes may emerge as markers because they change in the early stage of tumorigenesis in certain tumors although they may occur at the later stages of tumor progression in different systems (17). Activation of certain oncogenes (i.e., ras, myc, neu, jun, raf, or fos) has been associated with the development of lung cancer (17). However, the activation sequence of these genes, which would define which oncogenes are involved with the earliest stage of carcinogenesis, has not been well established. Other genes in this category include k-ras, c-myc, and certain specific chromosome regions. K-ras has been detected frequently in bronchial adenocarcinoma tumor tissue (18, 19) and has been associated with shortened survival in both early (20) and advanced stages of disease (21). Overexpression of the c-myc gene has been associated with growth dysregulation and loss of terminal differentiation in squamous cell (22, 23) and small cell tumors (24–26).

Losses of transcription factors or tumor suppressor genes may become markers of human head and neck and lung cancers. Probes that detect allelic deletion of specific chromosomal regions by restriction fragment length polymorphisms have frequently found loss of heterozygosity (expression of only a single allele) in SCLC on chromosomes 3p (100%), 13q (91%), and 17p (100%) (27). The karyotypes in NSCLC are very complex, but recurrent losses of 17p, 3p, and 11p (in 67%, 57%, and 48% of cases, respectively) suggest that these regions are “hot spots” for genetic alteration (28). Other candidate regions with breakpoints indicating that they are recessive oncogenes include 1q, 3q, 5p, 16q24, and 21p (29, 30).

Oncogene activation may alter the metabolic balance among cell growth, differentiation, and cell loss (17). By coding “transcription” protein, oncogenes activate other key genes to code for growth deregulation. The shift in the balance of cell differentiation to growth marks the selective clonal expansion that is characteristic of tumor proliferation. Two critical “early response” transcription factors, fos and jun, seem to be activated whenever mammalian cells respond to peptide growth factors (31). Bombesin, for example, a peptide growth factor released by pulmonary neuroendocrine cells, has been shown to induce growth and maturation of human fetal lung tissue in organ cultures (32). A functional membrane-associated bombesin receptor has recently been isolated from human SCLC carcinoma (NCI-H345) cells (33). This peptide has also been found in the
bronchial lavage fluid of asymptomatic chronic heavy smokers (34), suggesting that activation of certain oncogenes may occur by autocrine loop stimulation. Thus, markers of growth factor expression, insofar as they reflect activation of certain oncogenes, may hold promise for the early detection of lung cancer or premalignant lesions.

Since carcinogenesis begins with certain specific gene alterations and/or general genomic alterations that cause phenotypic changes such as dysregulation of proliferation and cell differentiation, we place biomarkers with potential utility in chemoprevention in five general classes (Table 1): (a) general genomic markers; (b) specific genetic markers; (c) proliferation markers; (d) differentiation markers; and (e) cell-loss markers. These biomarkers can reflect relatively early and site-specific carcinogenic changes; they also represent changes that are not site specific or that occur over extended periods of time.

### General Genomic Markers

**Micronuclei.** Micronuclei are chromosome or chromatid fragments formed in proliferating cells during clastogenic events such as DNA damage caused by carcinogens. The frequency of micronuclei was widely studied as a genomic marker in earlier human chemoprevention trials (35–40). Because micronuclei are easy to find and quantify, they are still the most studied of all potential biomarkers of intermediate end points. The micronuclei in the aerodigestive tract epithelium are formed in the proliferating basal cell layer, which gives rise to suprabasal cells that migrate to the epithelial surface and can eventually be detected in easily obtained exfoliated cells. The presence and frequency of micronuclei in tissue are believed to be quantitative reflections of ongoing DNA damage or genetic instability. A series of pilot trials has shown that high micronuclei frequency correlates with condemned-tissue cancer risk in individuals (e.g., smokers) who are at high risk for head and neck, lung, esophageal, or bladder carcinomas (36–39).

Because micronuclei frequency fits all four of the selection criteria already described, it seems probable that micronuclei would be a useful intermediate end point biomarker in trials of upper aerodigestive cancer chemoprevention. However, a critical analysis of the extensive data on this marker reveals problems with its use as the only marker (38). Stich et al. (37, 39–41) studied chemoprevention in high-risk groups from India and the Philippines who have remarkably intense and long-term carcinogenic exposure to betel nuts, uniform lifestyles and dietary habits, and consistently elevated oral micronuclei frequencies. These investigators reported that retinol and beta-carotene each suppressed micronuclei frequency in more than 90% of lesions after treatment; however, both the rates of clinical response to these agents and the rates of suppression of new lesions differed greatly (39–41). Therefore, this biomarker by itself may not be useful in screening or monitoring for active chemopreventive drugs. Another large placebo-controlled chemoprevention trial was conducted in subjects at high risk for esophageal cancer in Linxian County, People’s Republic of China (42). The investigators observed significant site-specific suppression of micronuclei (esophageal, not buccal) in subjects receiving chemopreventive agents (retinol, riboflavin, and zinc) (43). Again, the marker was not validated, since there was no significant reduction in the number of premalignant esophageal lesions after 1 year of chemopreventive intervention.

Despite these problems, studies of micronuclei as a biomarker have contributed to the early development of natural compounds, i.e., retinoids and carotenoids, as potential chemopreventive agents. Micronuclei frequency, however, indicates only an ongoing process of chromosomal damage, not accumulated genetic damage. Therefore, we are studying more specific alterations of genetic and phenotypic markers resulting from DNA damage by carcinogens.

**Chromosomal Alterations in Tumorigenesis.** Head and neck cancer provides a unique model system for the study of tumorigenesis and the development of biomarkers for several reasons. First, head and neck cancer probably represents a field cancerization process; the whole aerodigestive tract epithelium is repeatedly exposed to carcinogenic insult (e.g., tobacco and/or alcohol), placing the entire field at risk for tumor development (7, 44). The clinical evidence for this “field hypothesis” is the high frequency of multiple primary neoplasms in the aerodigestive tract and a higher-than-expected risk of synchronous and metachronous second primary tumors (10, 11). Second, head and neck cancer is thought to represent a multistep tumorigenesis process whereby a series of events may occur prior to tumor development (45). This is evinced by the presence of premalignant lesions adjacent to the tumor (46). Although these clinical and histologic findings support the notion of field cancerization and multistep tumorigenesis in the head and neck region, biomarkers for these processes are lacking. Therefore, we studied the genetic changes in the tissue at risk by using a multistep tumorigenesis model.

Although a variety of cytogenetic changes have been described for head and neck and lung cancers (46–48), the ability to develop a comprehensive list of specific chromosomal changes has been limited by impediments common to solid tumor cytogenetic studies, that is, the low frequency

---

**Table 1** Classes of biomarker candidates in upper aerodigestive tract tumors

<table>
<thead>
<tr>
<th>Class</th>
<th>Markers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genomic markers (general)</td>
<td>Nuclear alterations (e.g., micronuclei)</td>
</tr>
<tr>
<td></td>
<td>DNA content and flow cytometry</td>
</tr>
<tr>
<td>Chromosomal alterations</td>
<td></td>
</tr>
<tr>
<td>Specific genetic markers</td>
<td>Oncogene alterations [ras family, myc family, erb family (erbB1, erbB2/Her-2/neu)]</td>
</tr>
<tr>
<td></td>
<td>Src family (src, lck), retinoic acid receptors</td>
</tr>
<tr>
<td>Tumor suppressor genes [p53, retinoblastoma gene (rb), 3p (unidentified)]</td>
<td></td>
</tr>
<tr>
<td>Proliferation markers</td>
<td>Mitotic frequency (MPM-2)</td>
</tr>
<tr>
<td></td>
<td>Thymidine labeling index</td>
</tr>
<tr>
<td></td>
<td>Nuclear antigens (e.g., PCNA, DNA-polymerase alpha, Ki-67)</td>
</tr>
<tr>
<td></td>
<td>Polyamines, ornithine decarboxylase</td>
</tr>
<tr>
<td>Differentiation markers</td>
<td>Cytokeratins</td>
</tr>
<tr>
<td></td>
<td>Transthyretinase type I</td>
</tr>
<tr>
<td></td>
<td>Involutrin</td>
</tr>
<tr>
<td>Cell loss markers (apoptosis)</td>
<td>In situ end labeling of fragmented DNA</td>
</tr>
<tr>
<td></td>
<td>Bcl-2 expression</td>
</tr>
</tbody>
</table>
of mitotic figures from direct preparations, suboptimal chromosomal preparations, and multiple complexity of cytogenetic changes (49). Identification of karyotypic changes in premalignant lesions by conventional cytogenetic procedures is technically even more difficult than in tumors and has seldom been reported (50–52). Moreover, the spatial cellular distribution of genetic changes in premalignant and malignant lesions can not be defined by conventional cytogenetic techniques, since the tissue is disaggregated in the single cell preparations. Recently, ISH techniques have been developed that allow the direct detection of chromosomal abnormalities in interphase cells (53–57). This method has been applied to many types of solid tumors by using tumor cell lines or dissected tumor material (58–61). More recently, ISH was adapted for use on formalin-fixed, paraffin-embedded tissue sections by using nonisotopic, chromosome-specific DNA probes, and enzyme-mediated (e.g., peroxidase) immunohistochemical procedures (62–66). This technique allows direct visualization of chromosome changes in normal tissue, premalignant lesions, and tumor tissues with preservation of tissue architecture.

To visualize the accumulation of genetic alterations during head and neck tumorigenesis, and to determine the extent of the genetically altered field, we probed 25 squamous cell carcinomas of the head and neck and their adjacent premalignant lesions for numerical chromosome aberrations by nonisotopic ISH by using chromosome-specific centromeric DNA probes for chromosomes 7 and 17. Normal (control) oral epithelium from cancer-free nonsmokers showed to chromosome polysomy (i.e., cells with three or more chromosome copies), whereas histologically normal epithelium adjacent to tumors showed squamous cells with polysomies of chromosomes 7 and 17 (67). Moreover, the frequency of cells with polysomy increased as the tissues passed from histologically normal epithelium to hyperplasia to dysplasia to cancer. These chromosomal abnormalities in histologically normal and precancerous regions adjacent to the tumors support the concept of field carcinization, and the finding of progressive genetic changes as the tumor develops supports the concept of multistep tumorigenesis in the head and neck region (67).

This finding of chromosome number changes in the premalignant regions near head and neck tumors is not unexpected. Most of the carcinogens that cause upper aerodigestive tract tumors, including head and neck and lung carcinomas, are known to cause chromosomal abnormalities (68, 69). In fact, many other studies have reported increased frequencies of micronuclei at various sites in the aerodigestive tract, both in individuals exposed to various tobacco-related carcinogens and in those harboring premalignant lesions (37, 39–41). Individuals with head and neck cancer have also been shown to demonstrate increased in vitro sensitivity to chromosome-damaging agents, especially those individuals in whom second primary tumors develop (70, 71). This increased susceptibility to chromosome breakage might eventually lead to an accumulation of genetic alterations seen as chromosome polysomies in the heavily exposed epithelial tissue.

One of the goals of our study of head and neck carcinomas and their adjacent premalignant regions was to identify genetic biomarkers that might be useful for assessing risk of tumor development in the high-risk group and that might serve as intermediate end points in chemoprevention studies. The chromosome polysomies present even in histologically normal epithelium adjacent to tumors, and the increased frequency of chromosome polysomies as the tissues progressed to carcinomas suggest that the degree of generalized chromosome polysomies might be such a genetic biomarker. The advantage of such a biomarker is that it permits the sensitive detection of infrequent events that reflect accumulated genetic damage or genomic instability, events which are difficult to detect by bulk analysis (e.g., DNA content analysis). Detecting genomic instability is particularly important because it reflects an ongoing genetic process that translates to a higher risk.

The working hypothesis for our future studies could be that those individuals whose normal or premalignant epithelium exhibits the greatest degree of genetic abnormality might be expected to be at the highest risk for tumor development. Indeed, our preliminary retrospective studies in patients with oral premalignant lesions suggested that those individuals who exhibited greater than normal numbers of chromosomes were at the highest risk for the development of oral cancer (72). In this study, four of the seven patients with dysplastic lesions and one of the six patients with hyperplastic lesions showed evidence of chromosomal polysomies. More strikingly, of the four patients whose premalignant disease progressed to invasive cancer or carcinoma in situ, three of those patients exhibited chromosomal polysomies in more than 5% of cells, whereas only two of the nine patients who did not develop cancer reached this level of polysomies (72).

**Specific Genetic Markers**

**ras Gene Family.** A significant portion of human tumors from various sites in the body have been shown to contain activated oncogenes of the ras family (Harvey-ras, Kirsten-ras, and N-ras) (73–76). Oncogenes in the ras family are forms of the germline proto-oncogenes with specific point mutations that, when transfected onto NIH/3T3 murine fibroblasts, induce foci of morphologically altered cells (77–80). Normal ras genes code for proteins of molecular weights of approximately 21,000 that have guanine nucleotide-binding activity and are able to hydrolyze GTP (81). The proteins encoded by ras possess intrinsic GTPase activity which eventually leads to their inactivation, but this inactivation is greatly enhanced by a second protein, called the GTPase-activating protein (82). This protein has been shown to bind to the domain that is involved in the transduction of the ras gene signal, the “effector domain” of p21 (83, 84).

H-ras is found to be activated only infrequently, mainly in thyroid carcinomas (85), and N-ras activations are found predominantly in myeloproliferative disorders and in lymphomas (85, 86). For unknown reasons, K-ras is particularly associated with adenocarcinomas and has been reported to be activated in pancreatic cancers (87, 88), colorectal cancers (89, 90), and adenocarcinomas of the lung (91, 92). Rodenhuis et al. (93) reported that the majority of all ras mutations in lung cancer are found in adenocarcinomas, with frequencies of about 30% in smokers (41 of 141 samples) and about 5% in nonsmokers (2 of 40 samples). In contrast, SCLC is not associated with an activated ras oncogene (94). Activation of the K-ras gene has been reported to predict an unfavorable outcome: it identifies a subgroup of patients who have a very poor prognosis despite apparently successful surgery for stage I or II tumors (20). Another study by the National Cancer Institute essentially confirmed...
this observation by analyzing the K-ras gene in 52 cell lines derived from patients with lung adenocarcinomas (21).

Azuma et al. used immunohistochemical analysis of p21 expression in paraffin-embedded squamous cell head and neck carcinoma tissues, and found that the extent of expression of p21 was correlated with the degree of tumor differentiation, clinical stage, and clinical outcome (95). In this study, 59 of 121 tumor samples reacted to the monoclonal antibody (Y13-259) against p21 encoded by the v-ras gene of the Harvey murine sarcoma virus (96), whereas oral leukoplakia and normal mucosa samples did not react, indicating that they did not express this protein. Azuma et al. also reported that the expression of ras was associated with poor prognosis. The c-H-ras gene was also analyzed in 67 specimens of lymph node metastases and in 25 specimens of primary tumors obtained from 85 untreated patients with head and neck squamous cell carcinoma (97). Ten of 46 (22%) patients who were heterozygous for this locus had lost one allele. Polymerase chain reaction detected a mutation at codon 12 in only 2 of 54 (3.8%) tumors and no mutations at codon 61 (97).

Another study reported c-H-ras mutations at codon 12 in 2 of 37 squamous cell carcinomas of the head and neck and one of eight squamous carcinoma cell lines (98), indicating that the incidence of c-H-ras mutations is low in oral squamous cell carcinoma, at least among white caucasoid populations (99-101). It is interesting that oral carcinomas associated with chewing tobacco (quid) in an Indian population had a relatively high frequency of c-H-ras mutations, 20 of 57 (35%) cases, and all mutations were restricted to codons 61.2 (glutamine to arginine) and 12.2 (glycine to valine) (102). The striking difference in c-H-ras mutational frequencies in the two populations may have been due to any of several factors. For example, tobaccos used in India are likely to be from different strains or species of plant. Also, it is possible that the chewing of tobacco exposes the oral mucosa to concentrated levels of tobacco carcinogens for longer periods than does cigarette smoking. The question of genetic predisposition should also be considered. Finally, there is a possibility that the carcinogen(s) that induces c-H-ras mutations in oral mucosa may be present in some component of quid other than tobacco.

A few studies have been reported on ras gene mutations in premalignant lesions. K-ras mutations were found in adenomatous colon polyps from patients with familial polyposis coli, indicating that K-ras mutations could be a useful marker in colon tumorigenesis (103). In patients with colorectal tumors, K-ras gene mutations were detectable in DNA purified from stool specimens (104). These patients included both those with benign neoplasms (adenoma) and those with malignant neoplasms in the colonic epithelium. The c-H-ras mutations of premalignant lesions in the head and neck area have not been well explored. In particular, the premalignant lesions induced by the chewing of quid in Indian populations may be genetically altered before they develop to frank malignancy. Because of the relatively high incidence of c-H-ras mutations in oral squamous cell carcinomas in Indian populations, it is worthwhile to study these premalignant lesions to help in the design of a research strategy to prevent frank malignancy.

Activation of one of the ras genes may be useful for risk assessment during tumorigenesis in epithelial tissues. Further studies should be performed to determine whether expression of the ras gene or its mutations can be modulated by chemopreventive agents; if so, ras activation may also be used as an intermediate end point in chemoprevention trials.

**ErbB1, EGFR Gene.** The erbB1 oncogene was initially discovered as one of two oncogenes carried by the avian erythroblastosis virus (105). The corresponding proto-oncogene was found to encode a membrane-associated tyrosine kinase protein that was eventually identified as the receptor for EGF (106-108). On binding their respective ligands, the tyrosine kinase activity became stimulated several-fold, as indicated by enhanced autophosphorylation of the receptor, increased phosphorylation of exogenous substrates in vitro, and elevated phosphorylation of the tyrosine residues of several proteins in vivo (109). Two cell lines established from tumors of the head and neck area at different clinical stages were found to differ in the expression and in the tyrosine kinase activity of EGFR (109). The 1483 cells displayed a higher plating efficiency and clonogenicity in soft agar, suggesting that they have a more tumorigenic phenotype than the 183A cells. Analyses of EGFR levels by using R1 anti-EGFR serum indicated that the 1483 cells expressed 5-fold more receptors than the 183A cells. The autophosphorylation activity of both receptors was stimulated by addition of EGFR to isolated membrane preparations and intact cells, although the EGFR of the 1483 cells was much less responsive to EGF than that of the 183A cells (110).

The two cell lines having different characteristics, even though they originated from the same poorly differentiated squamous cell carcinomas of the head and neck, is an example of tumor heterogeneity. We can speculate that even one clonal cell with a high number of EGFR among a heterogenous population might eventually progress to frank malignancy. If these particular clonal cells were detected at the earlier stages of tumorigenesis, they would be good targets for chemopreventive therapy. If those malignant cells which had a high number of EGFR had increased growth and tumorigenic phenotype, we would expect the outcome of therapy, and the prognosis, to be poor.

To examine the possibility that EGFR expression could have predictive clinical value in head and neck squamous cell carcinoma, Santini et al. measured the EGFR levels in tumor tissues (111). In 59 of 60 samples, EGFR levels were higher in the tumor than in the corresponding normal controls. They also found a significant direct correlation between EGFR levels and tumor size and stage. Using immunohistochemical and cytometric techniques, expression of the Ki-67 antigen, EGFR, the TFR, and DNA ploidy were studied in 42 fresh samples of head and neck carcinomas. This study suggested that EGFR and TFR are widely distributed, especially on proliferating cells at the invading tumor margin. In addition, there is a close spatial correlation between cells that express EGFR and TFR and those that express Ki-67 antigen. Further follow-up is necessary to determine whether these parameters will be important prognostic values (112).

The EGFR gene has been found to be amplified in NSCLC (up to 20% in squamous cell types) (113-115), whereas the EGFR protein has been shown to be overexpressed in many NSCLC cells (approximately 90% of squamous cell types, 20-75% of adenocarcinomas, and infrequently in large cell or undifferentiated types) (114, 116, 117). However, EGFR amplification or protein overexpression has not been seen in SCLC cells. The overexpression of the EGFR protein may reflect the development of an autocrine growth loop, as EGF or transforming growth factor α...
is required by most epithelial cells, including NSCLC cells, for their growth (118). These findings have led us to develop clinical trials that use monoclonal antibodies against EGFR in the treatment (primary or adjuvant after surgery) in NSCLC (119, 120); these antibodies may also be useful in the prevention of lung cancer.

The expression of EGFR in premalignant lesions has not been well studied. To determine whether EGFR becomes overexpressed in premalignant lesions and to determine the consequences of such overexpression, we recently examined 36 head and neck squamous cell carcinomas with adjacent premalignant lesions and normal control epithelia. Using a monoclonal anti-EGFR antibody for immunohistochemical analysis of paraffin-embedded tissue sections, the degree of EGFR expression on epithelial cells was quantitated by computer-assisted image analysis (121). The level of EGFR expression was significantly higher in the adjacent normal tissue than in control samples (P = 0.021) that had never been exposed to tobacco and/or alcohol. These findings support the hypothesis of field cancerization (7). We also found another increment of EGFR expression, at the transition from the dysplastic lesions to squamous cell carcinomas, in two-thirds of the samples examined (P = 0.001). These preliminary results indicated that EGFR expression could be an important regulatory marker in the context of the multistep process of head and neck cancer development (121). To validate EGFR expression as a biomarker of an intermediate end point, we are currently exploring this marker in a large number of samples collected in chemoprevention trials.

**Int-2/Hst-1 Genes.** The hst-1 gene is one of the most frequently detected transforming genes after the ras gene family (122, 123). Because this gene encodes a protein that is homologous to a fibroblast growth factor and a int-2-encoded protein, it is assumed to be a member of the gene family that is involved in cell growth (124, 125). Both the hst-1 and int-2 genes are mapped to chromosome 11q13 (126), and their coamplification has been reported in bladder carcinoma (127), esophageal carcinoma (128), melanoma (129), gastric carcinoma (130), and breast carcinoma (130). At the same region of chromosome 11q13, other important genes (gst-π, PRAD, and cyclin D1) were both amplified and expressed. Coamplification of the hst-1 and int-2 genes in a hepatocellular carcinoma was accompanied by amplification of integrated hepatitis B virus DNA (131). The biological significance of coamplification of the hst-1 and int-2 genes is not clear.

The int-2 gene was amplified 3-fold to 5-fold in 5 of 10 laryngeal carcinomas and 2-fold to 3-fold in 5 of 11 carcinomas at other sites of the head and neck (132). However, the amplified int-2 gene has not yet shown any overexpression at either the mRNA or the protein level in head and neck squamous cell carcinomas. Adjacent histologically normal tissues from the same patients had only single copies of the gene. In a survey of head and neck tumor-derived cell lines, int-2 was amplified 9-fold in one, but not in three laryngeal cell lines (132). In another report, int-2 was found to be amplified in two of eight head and neck carcinomas (133). Although there was a suggestion that amplification of int-2 was proportionally correlated with tumor recurrence and clinical disease progression (132), a new study with a large patient sample is required to determine more precisely the significance of this gene amplification or overexpression in head and neck carcinomas.

To determine the chromosomal location of the amplified region and when during tumorigenesis the amplification process occurred, we examined head and neck squamous cell lines that were established by Sachs et al. (109) and paraffin-embedded tissue sections from which the cell lines were established, both by ISH, using both a cosmID probe for the int-2 gene and a biotin-labeled chromosome 11 painting probe. Three of 10 cell lines exhibited int-2 amplification, 2 of which (1386, 1986) were on chromosome 11 distal to the single copy gene; the third (886 cells) was on another chromosome. A fourth cell line (1486 cells), which did not have int-2 amplification, showed only a nonamplified single copy int-2 gene on chromosome 11. These findings correlated well with the amplification of int-2 as detected by Southern blotting. Paraffin blocks of the source tumors, which contained adjacent premalignant lesions, were analyzed with the int-2 probe to allow visualization of the timing of amplification during tumorigenesis. Two of the source tumors (1386 and 1986 cells) showed int-2 amplification in dysplastic and cancer regions, whereas a third (886 cells) showed amplification at the hyperplasia-to-dysplasia transition area and at the carcinoma in situ and tumor areas (134). These results suggest that int-2 amplification can occur in premalignant lesions prior to tumor development and validate our assumption that tumors develop through clonal evolution in situ in this model system (i.e., the changes persisted from premalignant to malignant cells). Therefore, int-2 amplification has the potential to be used as a marker for risk assessment and in chemoprevention trials (134).

**p53 Tumor Suppressor Gene.** The p53 gene, which encodes a nuclear protein, has been mapped to the short arm of chromosome 17 (17p13). The p53 protein is originally identified as a nuclear protein that bound to the large T antigen of the SV40 DNA tumor virus (135, 136). Although this gene was initially thought to act as a dominant oncogene, further investigation indicated that a mutant form existed (137). When the p53 gene was tested for its ability to transform cells, it was discovered that wild-type p53 gene could suppress transformation while the mutant form could induce transformation (137). However, many different types of alterations (rearrangement, deletion, insertion, or point mutation) have been observed to occur at different locations within the p53 gene in a wide variety of cell lines and human tumors (138–152).

The genetic alteration most frequently found in the p53 gene is a point mutation. Point mutation analyses have been confined primarily to exons 5 through 8, where the mutations are most frequently found in phylogenetically conserved regions. Hollstein et al. examined the frequency of nucleotide changes according to cancer type in a study of 280 p53 base substitutions (153). They observed some notable features: (a) most mutations in colorectal cancers, brain tumors, leukemia, and lymphomas occur at CpG dinucleotides, which are known to be mutational hot spots, and (b) G to A transitions constitute the majority of colon tumor mutations (31 of 39 mutations, 79%), whereas no G to T transitions were observed. More recent studies have detected C to T transversions in colon tumors (154, 155). In contrast, G to T transversions are the most frequent substitution in NSCLC (17 of 30 mutations, 57%) and liver tumors (14 of 19 mutations, 74%). In addition, the involvement of ultraviolet light in p53 mutations of squamous cell carcinoma of the skin was suggested by the presence in these...
cells of a CC to TT double-base change, which is known to be induced by ultraviolet light through the formation of pyrimidine photodimers (156). p53 mutational hotspots in human hepatocellular carcinoma vary according to the carcinogen by which they are induced (i.e., G to T transversion by hepatitis B virus, and G to C transversion by aflatoxin B1) (143). Thus, the specific p53 mutation sites may reflect the different carcinogen background of each tumor.

The p53 gene product is believed to function in cell cycle control (157–159). At least two stages in the cell cycle are regulated in response to DNA damage, the G1-S and the G2-M transitions. These transitions serve as checkpoints at which cell cycle progression is delayed to allow repair of DNA damage before the cell enters either S phase, when DNA damage would result from DNA adduct formation leading to mutation or chromosomal damage, or M phase, when chromosome breaks would cause the loss of genomic material from daughter cells. Checkpoints are believed to be surveillance mechanisms that can detect DNA damage and active signal transduction pathways and then regulate replication or segregation machinery and, possibly, repair activities (160).

Since there are p53 mutants that do not exhibit the cell cycle arrest or delay that occurs in wild-type p53 in response to DNA damage, both the G1-S and G2-M checkpoints are known to be under genetic control. There is strong evidence that p53 inhibits the G1-S transition, because high levels of p53 block cell cycle progression at the G1-S checkpoint (157). Tumor cells that lack p53 or that have dominant mutant forms of p53 lack the G1-S delay that occurs on exposure to ionizing radiation (161). The loss of a G1 checkpoint in mammalian cells is not associated with increased sensitivity to the lethal effects of ionizing radiation (162), but it is associated with an increase in mutational frequency. The G2-M checkpoint is abolished by mutation of p53 in a number of yeast genes (163) and by treatment of mammalian cells with caffeine (164). The epithelium of the head and neck area is constantly exposed to carcinogens, and p53 may affect how these damaged cells respond to this insult. We speculate that the upregulation of p53 in response to damage induced by carcinogens increases its susceptibility to mutation. There are human lymphoma cell lines that have increased sensitivity to irradiation while the cells are arrested during S phase (165). Therefore, p53 mutation is eventually related to cell proliferation.

The normal p53 protein has a very short half-life (6–20 min), whereas the mutant form has a half-life of up to 6 h. It has been inferred, therefore, that the presence of detectable p53 protein implies mutation (166). Iggo et al. (167) found increased p53 oncoprotein staining in the tissues of lung cancers that are associated with smoking. They reported elevated p53 protein levels in 14 of 17 (82%) squamous cell carcinomas and in only eight of 21 (38%) nonsquamous cell carcinomas. Similarly, Chiba et al. (144) reported finding p53 mutations in 65% of the lung squamous cell carcinomas and in 36% of the nonsquamous tumors. The association of smoking with squamous cell carcinomas of the lung provides further evidence for a link between p53 mutations and smoking. In a similar study of p53 in SCLC cell lines, D’Amico et al. (168) found that 100% of the SCLC cells had p53 mutations.

More recently, inactivating mutations in the p53 gene have been identified in truly preneoplastic lesions, namely, Barrett's esophagus (169), a precursor to adenocarcinomas of the esophagus. In bronchial epithelium, the p53 protein was detected in 0% of normal mucosa, 6.7% of squamous metaplasia, 29.5% of mild dysplasia, 59.2% of severe dysplasia, and 58.5% of carcinomas in situ (170). Nees et al. (171) studied p53 mutations in the respiratory epithelium either adjacent to or at significant distance from primary head and neck tumors. They observed p53 mutations in the distant epithelia of these patients and concluded that mutation of p53 is an early event in head and neck carcinogenesis, supporting the field carcinogenesis hypothesis. A similar observation was also made by Boyle et al. (172).

We studied p53 protein expression in 33 patients with head and neck squamous cell carcinomas whose tissue sections contained adjacent normal epithelium, hyperplasia, and/or dysplasia. Fifteen of 33 (45%) head and neck tumor specimens expressed p53, but none of the normal controls (tissue specimens from cancer-free nonsmokers) expressed detectable p53 protein. However, 5 of 24 (21%) specimens of normal epithelium adjacent to tumors, seven of 24 (29%) hyperplastic lesions, and nine of 20 (45%) dysplastic lesions expressed p53 (173). We conclude that p53 expression can be altered in very early phases of head and neck tumorigenesis. To determine whether increased expression of p53 was associated with a gene mutation, we performed a combination of polymerase chain reaction-single-strand conformation polymorphism and direct genomic sequencing on one representative case that expressed a high level of p53 in the tumor but not in its adjacent normal epithelium. We found that the adjacent normal epithelium had a wild-type p53 at codon 174 in exon 5. In the tumor sample, however, 10 base pairs had been deleted at this position. This finding confirmed that there was a good correlation between high levels of p53 expression and mutation. Further work is needed, however, to establish the relationship between low levels of p53 expression and gene mutations. We also evaluated p53 protein expression in 27 premalignant oral lesions and in normal oral mucosa of eight healthy nonsmoking control individuals. In 14 of the 27 lesions (52%), p53 was expressed in more than 5% of cells calculated, while there was no p53 expression in nonsmoking controls. Eight lesions (all in current or former smokers) had very high baseline levels of p53. The degree of response to 3 months of treatment with high-dose 13-cis retinoic acid correlated inversely with baseline p53 expression level (174). This preliminary study suggests that p53 expression may be an excellent predictor of risk and may serve as an intermediate biomarker in chemoprevention trials (173, 174).

Proliferation Markers
It is hypothesized that only those cells with high proliferative activity could be associated with premalignant and malignant tissue changes during tumorigenesis. Lee et al. (175) initially used antibodies against the cell proliferation markers, including PCNA, and found PCNA to be quite useful as a marker for proliferating cells, even in routinely processed formalin-fixed, paraffin-embedded tissue sections (176). PCNA is a Mr 36,000 acidic, nonhistone, nucleolar protein whose expression is associated with the late G2 phases of the cell cycle (177). It is an auxiliary protein to DNA polymerase δ, which has a 261-amino-acid polypeptide with high aspartic and glutamic acid contents and plays a critical role in the initiation of cell proliferation (178, 179).
To better understand the relationship between PCNA expression in tissue and proliferation status, specimens of head and neck and colorectal cancers were excised from patients after infusion of BrdUrd (180). The specimens were embedded in paraffin and examined. Adjacent tumor sections were analyzed for PCNA expression and for incorporated BrdUrd (using anti-PCNA antibody and anti-BrdUrd antibody, respectively). Regions of tumors that were high in PCNA-positive cells also had high BrdUrd uptake, and vice versa. In all cases, the proportion of PCNA-positive cells was higher than the proportion of BrdUrd-positive cells. These results were not surprising, since PCNA is expressed in most proliferating phases (G1, S, and G2), whereas BrdUrd uptake marks only the S phase (179).

With the hypothesis that PCNA expression is dysregulated in tumors, 107 NSCLC tissue sections were examined for PCNA expression (175). Squamous cell carcinomas showed the highest proliferative activity, with a mean of 40% PCNA-positive cells (range, 2–90%); adenocarcinomas had a mean of only 5% PCNA-positive cells (range, 0–70%), and large cell carcinomas had a mean of 15% (range, 3–80%). The PCNA-positive fraction became progressively larger in areas of squamous metaplasia and carcinoma in situ (175). To better understand tumorigenesis in head and neck cancer, we studied 33 formalin-fixed, paraffin-embedded tissue specimens from five different sites of head and neck squamous cell carcinomas that contained adjacent normal epithelium, hyperplasia, and/or dysplasia (181). PCNA expression was assessed by semiquantitative scoring in three epithelial layers (basal, parabasal, and suprabasal). The labeling index (the number of positively stained cells divided by the total number of cells counted) and the weighted mean index of PCNA expression (the sum of the number of counted cells multiplied by degree of intensity (0–3) in each cell and divided by the total number of counted cells) were calculated to represent the level of PCNA expression. What was interesting was that normal epithelium adjacent to the tumor had much more proliferative activity than control epithelium (from cancer-free nonsmokers). Furthermore, PCNA expression increased as the tissue progressed from adjacent normal epithelium to hyperplasia (P < 0.001), hyperplasia to dysplasia (P < 0.001), and dysplasia to squamous cell carcinomas (P = 0.065); the total increase in PCNA expression from adjacent normal epithelium to squamous cell carcinomas ranged from 4-fold to 10-fold (181). As the tissue progressed to carcinoma, we observed not only increases in the number of proliferating cells but also in the amount of PCNA expressed by each labeled cell. These studies indicate that PCNA could be a useful biomarker for multistep carcinogenesis in head and neck cancer and that its expression could serve as an intermediate endpoint in chemoprevention trials (181).

As part of an ongoing chemoprevention trial in which chronic smokers (≥15 pack-years) were screened for squamous metaplasia before being randomized to receive either 13-cis retinoic acid or placebo, our group examined PCNA expression in bronchial biopsy sections obtained from six standardized sites in the major bronchial trees (180). In this study, 165 samples were evaluated for PCNA expression and histologic status. Among the 81 biopsied specimens showing histologically normal epithelium, only 12% had more than 1% PCNA-positive cells, and no specimen had more than 5% positive cells. In contrast, 37% (19 of 52) of the hyperplasia specimens and 50% (10 of 20) of the metaplasia specimens had more than 1% PCNA-positive cells. Among those biopsied specimens showing dysplasia with metaplasia, 58% (7 of 12) had more than 1% PCNA-positive cells. These results suggest a significant correlation between increases in proliferative activity and histological progression in epithelium at high risk of tumor development (180). In a study of esophageal premalignancy by Yang et al. (182) and a study of subjects at high risk for colon cancer by Lipkin et al. (183, 184), the patterns of expression of the proliferation marker and tritiated thymidine incorporation were similar to the patterns of PCNA expression we observed in the lung and head and neck cancers. These two pilot studies also suggested that the chemopreventive drug can suppress this marker. A large scale of chemoprevention study incorporating studies of biomarkers for oral premalignancy and second primary cancer prevention is ongoing at M. D. Anderson Cancer Center. We hope we will have more definite answers to these questions in the near future.

**Squamous Cell Differentiation Markers**

Upper aerodigestive epithelia differentiate along the squamous pathway in carcinogenesis, and understanding of this process is important for developing chemoprevention strategies (185, 186). To establish a preclinical carcinogenesis model and to determine which markers are important, we examined several markers, including cytokeratin and TGase 1, in the DMBA-induced hamster buccal pouch model (187, 188). We applied DMBA (0.5%) in heavy mineral oil to the hamster buccal pouch three times per week for up to 16 weeks. TGase 1 was expressed at a limited level in normal buccal mucosa, at a low level in the basal layer of hyperplastic lesions, and at a somewhat higher level in dysplasia; its expression was markedly increased in squamous cell carcinoma (187). The cytokeratin assayed in this study were K14 (M, 55,000), K1 (M, 67,000), and K13 (M, 47,000). Normal hamster cheek pouch epithelium expressed K14 in the basal layer and K13 in the suprabasal and differentiated layers; K1 was not detected. In hyperplasia, K14 was no longer restricted to the basal layer but was expressed in differentiated cells. The same pattern was observed in dysplasia; K14 expression was in squamous cell carcinomas. However, K13 was preserved in hyperplastic epithelium during all stages of carcinogenesis, including anaplastic or differentiated areas. Expression of K1, in contrast, started as a weak and patchy pattern after 2 weeks of DMBA treatment but became stronger and more homogeneous at 8 weeks of treatment. However, K1 was almost absent in squamous cell carcinomas. We concluded that the pattern of keratin expression could be an important tool in the study of carcinogenesis (188).

Another marker for squamous cell differentiation is involucrin, one of the major protein components of cornified envelopes (189). This protein undergoes extensive cross-linking by the membrane-associated (particulate) enzyme type 1 TGase, which catalyzes the formation of ϵ-(r-glutamyl isopeptide) linkages between protein-bound glutamine residues and primary amines such as protein-bound lysine (190, 191). These proteins are expressed in the upper stratum spinosum and corneal layers of the epidermis (191). Squamous differentiation of keratinocytes is usually accompanied by increases in the levels of involucrin and TGase 1, but the expression of involucrin precedes the expression of TGase 1 (189, 190). Involucrin is expressed in premalignant lesions and squamous cell carcinomas (192,
This strategy differs from that of large-scale chemoprevention trials in the upper aerodigestive tract to reduce the incidence of cancer development in that field of tissue. As described previously, however, the major obstacle to such trials is that the study end point (i.e., cancer) generally takes years to become detectable. It would be ideal to define intermediate end points that reflect whether chemopreventive agents have an effect on the tissue at risk (4, 5) and, if so, their mechanisms of action at the tissue level. First, biomarkers would be useful for assessing the risk of tumor development in high-risk tissue and in premalignant lesions. Second, these markers would allow us to better understand the pathobiology of clinical response to the chemopreventive treatment. Third, such markers would be good indicators of intermediate end points in clinical chemoprevention trials, allowing prediction of patients’ responsiveness before the final end point, cancer, could be reached. In such trials, we would want to know whether clinical outcome of the treatment is due to (a) reversal of the abnormal clones at the genetic level; (b) phenotypic reversal of the abnormal clones; or (c) partial suppression of less-affected clones (if two or more distinct clones can be identified in the lesions). These questions could be answered by examining both genetic markers and phenotypic markers on the tissue samples obtained before and after chemopreventive therapy. Finally, although strict validation of any biomarker as a true intermediate end point of cancer development may take many years of follow-up in large-scale clinical trials, current biomarker candidates are an important adjunct to the development of new chemopreventive agents and to the rational design of future intervention trials.

In this review, we described several candidates for genetic and phenotypic biomarkers, all of which need to be validated in clinical trials. In the future, validated comprehensive panels of biomarkers, indicating early and intermediate stages of the multistep carcinogenesis process, may provide new standard end points and even replace cancer incidence as the sole end point for chemoprevention trials. If this occurs, an early step in preventive studies will be to determine whether the agents have any activity at all in affecting the tumorigenesis process. Later on, the active agents can be evaluated for their effects on tumor formation. This strategy differs from that of large-scale chemopreventive studies; it requires a smaller number of subjects and may be conducted within a relatively short period of time.

References


intronic point mutation in colorectal tumors. Biochem. Biophys. Res. Com-
157. Kastan, M. B., Zhan, Q., El-Deiry, W. S., Carris, F., Jacks, T., Welsh, W. K., Plunkett, B. S., Vogelstein, B., and Forman, A. J. A mammalian cell cycle checkpoint pathway utilizing p53 and CAD2D5 is defective in ataxia-
Biomarkers in upper aerodigestive tract tumorigenesis: a review.

D M Shin, W N Hittelman and W K Hong